UNITED STATES PATENT AND TRADEMARK OFFICE

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
09/909,857	07/23/2001	Evan Hildreth	12121-002001	8772

26171   7590   08/23/2004

FISH & RICHARDSON P.C.
1425 K STREET, N.W.
11TH FLOOR
WASHINGTON, DC 20005-3500

EXAMINER
DHARIA, PRABODH M

ART UNIT	PAPER NUMBER
2673	

DATE MAILED: 08/23/2004

Please find below and/or attached an Office communication concerning this application or proceeding.

PTO-90C (Rev. 10/03)

	Application No.		Applicant(s)	
Office Action Summary	09/909,857		HILDRETH ET AL.	
	Examiner		**Art Unit**	
	Prabodh M Dharia		2673	

-- *The MAILING DATE of this communication appears on the cover sheet with the correspondence address* --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE _3_ MONTH(S) FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If the period for reply specified above is less than thirty (30) days, a reply within the statutory minimum of thirty (30) days will be considered timely.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on _23 July 2001_.

2a)☐ This action is **FINAL**.   2b)☒ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) _1-98_ is/are pending in the application.

    4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) _1-98_ is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☒ The drawing(s) filed on _23 July 2001_ is/are: a)☒ accepted or b)☐ objected to by the Examiner.

    Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

    Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

    a)☐ All b)☐ Some * c)☐ None of:

      1.☐ Certified copies of the priority documents have been received.

      2.☐ Certified copies of the priority documents have been received in Application No. _____.

      3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

    * See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☒ Notice of References Cited (PTO-892)

2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)

3) ☒ Information Disclosure Statement(s) (PTO-1449 or PTO/SB/08) Paper No(s)/Mail Date _5.7_.

4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____ .

5) ☐ Notice of Informal Patent Application (PTO-152)

6) ☐ Other: _____ .

Priority

1.   It is noted that this application appears to claim subject matter disclosed in prior

Application No. 60/220/223, filed 07/24/2000. A reference to the prior application has been

inserted as the first sentence of the specification of this application and in an application data

sheet (37 CFR 1.76), as applicant intends to rely on the filing date of the prior application under

35 U.S.C. 119(e).

Claim Rejections - 35 USC § 102

2.   The following is a quotation of the appropriate paragraphs of 35 U.S.C. 102 that form the

basis for the rejections under this section made in this Office action:

> A person shall be entitled to a patent unless –
>
> (e) the invention was described in (1) an application for patent, published under section 122(b), by another filed in the United States before the invention by the applicant for patent or (2) a patent granted on an application for patent by another filed in the United States before the invention by the applicant for patent, except that an international application filed under the treaty defined in section 351(a) shall have the effects for purposes of this subsection of an application filed in the United States only if the international application designated the United States and was published under Article 21(2) of such treaty in the English language.

3.   Claims 1-12, 23,54-70 are rejected under 35 U.S.C. 102(e) as being anticipated by Pryor

et al. (US 2004/0046736 A1).

Regarding Claim 1, Pryor et al. teaches a method of using stereo vision to interface with a

computer (page 1, paragraph 3, Lines 4-7), the method comprising: capturing a stereo image

(page 8, paragraph 173, Lines 1-3, page 10, paragraph 239, Lines 1-4); processing the stereo

image to determine position information of an object in the stereo image (page 8, paragraph 170,

Lines 1-7, page 10, paragraph 239, Lines 5,6), the object being controlled by a user (page 10,

paragraph 239); and using the position information to allow the user to interact with a computer

application (page 10, paragraph 239-241).

Regarding Claim 2, Pryor et al. teaches the step of capturing the stereo image further

includes capturing the stereo image using a stereo camera (page 1, paragraph 3, Lines 4-7).

Regarding Claim 3, Pryor et al. teaches recognizing a gesture associated with the object

by analyzing changes in the position information of the object, and controlling the computer

application based on the recognized gesture (page 10, paragraph 239-241).

Regarding Claim 4, Pryor et al. teaches determining an application state of the computer

application; and using the application state in recognizing the gesture (Page 10, paragraph 239-

241).

Regarding Claim 5, Pryor et al. teaches the object is the user (page 10, paragraph 243,

244).

Regarding Claim 6, Pryor et al. teaches the object is a part of the user (page 10, paragraph

243,244).

Regarding Claim 7, Pryor et al. teaches providing feedback to the user relative to the

computer application (page 10, paragraph 238-244)

Regarding Claim 8, Pryor et al. teaches processing the stereo image to determine position information of the object further includes mapping the position information from position coordinates associated with the object to screen coordinates associated with the computer application (page 10, paragraphs 238-244).

Regarding Claim 9, Pryor et al. teaches processing the stereo image further includes processing the stereo image to identify feature information and produce a scene description from the feature information (page 8, paragraphs 170-173).

Regarding Claim 10, Pryor et al. teaches analyzing the scene description in a scene analysis process to determine position information of the object (page 8, paragraph 170-173)

Regarding Claim 11, Pryor et al. teaches processing the stereo image further includes: analyzing the scene description to identify a change in position of the object; and mapping the change in position of the object (*page 9, paragraph 224)

Regarding Claim 12, Pryor et al. teaches processing the stereo image to produce the scene description further includes: processing the stereo image to identify matching pairs of features in the stereo image; and calculating a disparity and a position for each matching feature pair to create a scene description (page 9, paragraph 224).

Regarding Claim 23, Pryor et al. teaches for each feature pair in the scene description, calculating real world coordinates by transforming the disparity and position of each feature pair relative to the real world coordinates of the stereo image (page 10, paragraph 238-241).

Regarding Claim 54, Pryor et al. teaches a stereo vision system (page 1, paragraph 3, Lines 5-7, paragraph 28, 11-13) for interfacing with an application program running on a computer (page 1, paragraph 3, Lines 5-7, page 5, paragraph 111, Line 2) the stereo vision system (page 1, paragraph 3, Lines 5-7, paragraph 28, 11-13) comprising: first and second video cameras (page 5, paragraph 111, Lines 2,3) arranged in an adjacent configuration and operable to produce a series of stereo video images (page 6, paragraph 119, Lines 1-7, figure 1c); and a processor operable to receive the series of stereo video images and detect objects appearing in an intersecting field of view of the cameras (page 6, paragraphs 119-121), the processor executing a process to: define an object detection region in three-dimensional coordinates relative to a position of the first and second video cameras (page 11, paragraphs 247, 249); select a control object appearing within the object detection region; and map position coordinates of the control object to a position indicator associated with the application program as the control object moves within the object detection region (page 11, paragraphs 247-249).

Regarding Claim 55, Pryor et al. teaches the process selects as a control object a detected object appearing closest to the video cameras and within the object detection region (page 11, paragraphs 247-249).

Regarding Claim 56, Pryor et al. teaches the control object is a human hand (page 11, paragraph 256).

Regarding Claim 57, Pryor et al. teaches a horizontal position of the control object relative to the video cameras is mapped to an x-axis screen coordinate of the position indicator (page 10, paragraph 238).

Regarding Claim 58, Pryor et al. teaches a vertical position of the control object relative to the video cameras is mapped to a y-axis screen coordinate of the position indicator (page 10, paragraph 238).

Regarding Claim 59, Pryor et al. teaches the processor is configured to: map a horizontal position of the control object relative to the video cameras to a x-axis screen coordinate of the position indicator; map a vertical position of the control object relative to the video cameras to a y-axis screen coordinate of the position indicator; and emulate a mouse function using the combined x-axis and y-axis screen coordinates provided to the application program (page 10, paragraphs 238-242).

Regarding Claim 60, Pryor et al. teaches the processor is further configured to emulate buttons of a mouse using gestures derived from the motion of the object position (page 10, paragraphs 241-244)

Regarding Claim 61, Pryor et al. teaches the processor is further configured to emulate buttons of a mouse based upon a sustained position of the control object in any position within the object detection region for a predetermined time period (page 10, paragraphs 241-244).

Regarding Claim 62, Pryor et al. teaches the processor is further configured to emulate buttons of a mouse based upon a position of the position indicator being sustained within the bounds of an interactive display region for a predetermined time period (page 10, paragraphs 241-244).

Regarding Claim 63, Pryor et al. teaches the processor is further configured to map a z-axis depth position of the control object relative to the video cameras to a virtual z-axis screen coordinate of the position indicator (page 27, paragraph 529).

Regarding Claim 64, Pryor et al. teaches the processor is further configured to: map a x-axis position of the control object relative to the video cameras to an x-axis screen coordinate of the position indicator; map a y-axis position of the control object relative to the video cameras to a y-axis screen coordinate of the position indicator; and map a z-axis depth position of the control object relative to the video cameras to a virtual z-axis screen coordinate of the position indicator (page 10, paragraphs 229-242, page 13, paragraph 297)

Regarding Claim 65, Pryor et al. teaches a position of the position indicator being within the bounds of an interactive display region triggers an action within the application program

(page 12, paragraphs 262-272).

Regarding Claim 66, Pryor et al. teaches movement of the control object along a z-axis

depth position that covers a predetermined distance within a predetermined time period triggers a

selection action within the application program (page 10, paragraphs 229-242, page 13,

paragraph 297).

Regarding Claim 67, Pryor et al. teaches a position of the control object being sustained

in any position within the object detection region for a predetermined time period triggers a

selection action within the application program (page 12, paragraphs 262-272, page 10,

paragraphs 229-242, page 13, paragraph 297).

Regarding Claim 68, Pryor et al. teaches a stereo vision system (page 1, paragraph 3,

Lines 5-7, paragraph 28, 11-13) for interfacing with an application program running on a

computer (page 1, paragraph 3, Lines 5-7, page 5, paragraph 111, Line 2) the stereo vision

system (page 1, paragraph 3, Lines 5-7, paragraph 28, 11-13) comprising: first and second video

cameras (page 5, paragraph 111, Lines 2,3) arranged in an adjacent configuration and operable to

produce a series of stereo video images (page 6, paragraph 119, Lines 1-7, figure 1c); and a

processor operable to receive the series of stereo video images and detect objects appearing in an

intersecting field of view of the cameras (page 6, paragraphs 119-121), the processor executing a

process to: define an object detection region in three-dimensional coordinates relative to a

position of the first and second video cameras (page 11, paragraphs 247, 249); select a control

object appearing within the object detection region; and map position coordinates of the control

object to a position indicator associated with the application program as the control object moves

within the object detection region (page 11, paragraphs 247-249) and define sub regions within

the object detection region; identify a sub region occupied by the control object; associate with

that sub region an action that is activated when the control object occupies that sub region; and

apply the action to interface with a computer application (page 12, paragraphs 273-276, page 13,

paragraphs 291-294, page 10, paragraphs 230 –239, sub-regions rotated or targeted to blend

together or access to alter original target in sub-region with to generate 3D image).

Regarding Claim 69, Pryor et al. teaches the action associated with the sub region is

further defined to be an emulation of the activation of keys associated with a computer keyboard

(page 12, paragraphs 273-276).

Regarding Claim 70, Pryor et al. teaches a position of the control object being sustained

in any sub region for a predetermined time period triggers the action (page 12, paragraphs 262-

276, page 10, paragraphs 229-242, page 13, paragraph 297).

Claim Rejections - 35 USC § 103

4.   The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all

obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in
> section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are
> such that the subject matter as a whole would have been obvious at the time the invention was made to a person
> having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the
> manner in which the invention was made.

5.   Claims 13-22,24-53 are rejected under 35 U.S.C. 103(a) as being unpatentable over Pryor

et al. (US 2004/0046736 A1) as applied to claims 1-12,23, 54-70 above, and further in view of

Gordon et al. (6,661,918 B1) and Onda (6,125,198).

Regarding Claim 13, Pryor et al. teaches capturing the stereo image further includes

capturing a reference image from a reference camera and a comparison image from a comparison

camera; and processing the stereo image further includes processing the reference image and the

comparison image to create pairs of features (page 6, paragraph 117, Lines 1-4, paragraph 119,

paragraph 121, 128, page 7, paragraph 134, Lines 3-5, paragraph 135, 136, 148-160, page 10,

paragraph 238,239).

However, Prior et al. fails to specifically recite a reference camera.

However, Gordon et al. teaches a reference camera (Col. 4, Lines 21-25).

Thus it would have been obvious to one in the ordinary skill in the art at the time of

invention was made to incorporate the teaching of Gordon et al. in Pryor et al. teaching, to be

able to provide computer and a technique for automatically distinguishing between a

background scene and foreground objects in an image.

Pryor et al. teaches capturing the stereo image further includes capturing a reference

image from a reference camera and a comparison image from a comparison camera; and

processing the stereo image further includes processing the reference image and the comparison

image to create pairs of features (page 6, paragraph 117, Lines 1-4, paragraph 119, paragraph

121, 128, page 7, paragraph 134, Lines 3-5, paragraph 135, 136, 148-160, page 10, paragraph

238,239).

However, Prior et al. fails to specifically recite capturing a reference image from a camera.

However, Onda teaches capturing a reference image from a camera (Col. 11, Lines 9-25, Lines 43-67).

Thus it would have been obvious to one in the ordinary skill in the art at the time of invention was made to incorporate the teaching of Onda in Pryor et al. teaching, to be able to provide computer and a technique for matching stereo images and a method of detecting disparity between these images to detect positional information in the image pickup space based on stereo images, volume compression of overall stereo images display control of these stereo images and for the optical flow extraction of moving images.

Regarding Claim 14, Onda teaches processing the stereo image to identify matching pairs of features in the stereo image further includes: identifying features in the reference image; generating for each feature in the reference image a set of candidate matching features in the comparison image; and producing a feature pair by selecting a best matching feature from the set of candidate matching features for each feature in the reference image (Col. 11, Lines 9-25, Lines 43-58).

Regarding Claim 15, Pryor et al. teaches processing the stereo image further includes filtering the reference image and the comparison image (page 6, paragraph 117, Lines 1-4, paragraph 119, paragraph 121, 128, page 7, paragraph 134, Lines 3-5, paragraph 135, 136, 148-160, page 10, paragraph 238,239).

Gordon et al. teaches processing the stereo image further includes filtering the reference image and the comparison image (Col. 5, Line 43, to Col. 6, Line 3).

Regarding Claim 16, Pryor et al. teaches the feature pair further includes: calculating a match score and rank for each of the candidate matching features; and selecting the candidate matching feature with the highest match score to produce the feature pair (page 13, paragraph 293-295).

Regarding Claim 17, Pryor et al. teaches generating for each feature in the reference image, a set of candidate matching features further includes; selecting candidate matching features from a predefined range in the comparison image (page 31, paragraph 293-295)

Regarding Claim 18, Pryor et al. teaches feature pairs are eliminated based upon the match score of the candidate matching feature (page 13, paragraph 293-295).

Regarding Claim 19, Pryor et al. teaches feature pairs are eliminated if the match score of the top ranking candidate matching feature is below a predefined threshold (page 7, paragraph 135-160, page 10, paragraph 238,239, page 13, paragraph 293-295).

Regarding Claim 20, Pryor et al. teaches the feature pair is eliminated if the match score of the top ranking candidate matching feature is within a predefined threshold of the match score of a lower ranking candidate matching feature (page 7, paragraph 135-160, page 10, paragraph

238,239, page 13, paragraph 293-295).

Regarding Claim 21, Pryor et al. teaches calculating the match score further includes:

identifying those feature pairs that are neighboring; adjusting the match score of feature pairs in

proportion to the match score of neighboring candidate matching features at similar disparity;

and selecting the candidate matching feature with the highest adjusted match score to create the

feature pair (page 6, paragraph 117, Lines 1-4, paragraph 119, paragraph 121, 128, page 7,

paragraph 134, Lines 3-5, paragraph 135-160, page 10, paragraph 238,239, page 13, paragraph

293-295).

Onda teaches calculating the match score further includes: identifying those feature pairs

that are neighboring; adjusting the match score of feature pairs in proportion to the match score

of neighboring candidate matching features at similar disparity; and selecting the candidate

matching feature with the highest adjusted match score to create the feature pair (Col. 11, Lines

9-25, Lines 43-58, Col. 14, Lines 29-59, Col. 15, Lines 13-39).

Regarding Claim 22, Pryor et al. teaches feature pairs are eliminated by: applying the

comparison image as the reference image and the reference image as the comparison image to

produce a second set of feature pairs; and eliminating those feature pairs in the original set of

feature pairs which do not have a corresponding feature pair in the second set of feature pairs

(page 6, paragraph 117, Lines 1-4, paragraph 119, paragraph 121, 128, page 7, paragraph 134,

Lines 3-5, paragraph 135-160, page 10, paragraph 238,239).

Regarding Claim 24, Pryor et al. teaches selecting features further includes dividing the

reference image and the comparison image of the stereo image into blocks (page 21, paragraph

451 to page 22, paragraph 451-454).

Regarding Claim 25, Pryor et al. teaches the feature is described by a pattern of

luminance of the pixels contained with the blocks (page 8, paragraph 173-176, page 21,

paragraph 451 to page 22, paragraph 451-454).

Regarding Claim 26, Pryor et al. teaches dividing further includes dividing the images

into pixel blocks having a fixed size (page 8, paragraph 173-176, page 21, paragraph 451 to page

22, paragraph 451-454).

Regarding Claim 27, Onda teaches the pixel blocks are 8.times.8 pixel blocks (Col. 13,

Line 60 to Col. 14, Line 28).

Regarding Claim 28, Onda teaches analyzing the scene description to determine the

position information of the object further includes cropping the scene description to exclude

feature information lying outside of a region of interest in a field of view (Col. 11, Lines 9-25,

Lines 43-58).

Regarding Claim 29, Pryor et al. teaches cropping further includes establishing a

boundary of the region of interest (page 13, paragraph 293,294).

Regarding Claim 30, Gordon et al. teaches analyzing the scene description to determine

the position information of the object further includes: clustering the feature information in a

region of interest into clusters having a collection of features by comparison to neighboring

feature information within a predefined range; and calculating a position for each of the clusters

(Col. 5, Line 43 to Col. 6, line 3, Col. 11, Line 29 to Col. 12, Line 6).

Onda teaches analyzing the scene description to determine the position information of the

object further includes: clustering the feature information in a region of interest into clusters

having a collection of features by comparison to neighboring feature information within a

predefined range; and calculating a position for each of the clusters (Col. 11, Lines 9-25, Lines

43-58, Col. 13, Line 60 to Col. 14, Line 59).

Regarding Claim 31, Grdon et al. teaches eliminating those clusters having less than a

predefined threshold of features (Col. 5, Line 43 to Col. 6, line 3, Col. 11, Line 29 to Col. 12,

Line 6).

Onda teaches eliminating those clusters having less than a predefined threshold of

features (page 10, paragraph 234, Lines 1-4, paragraph 238).

Regarding Claim 32, Pryor et al. teaches selecting the position of the clusters that match a

predefined criteria; recording the position of the clusters that match the predefined criteria as

object position coordinates; and outputting the object position coordinates (page 10, paragraph 234, Lines 1-4, paragraph 238).

Gordon et al. teaches selecting the position of the clusters that match a predefined criteria; recording the position of the clusters that match the predefined criteria as object position coordinates; and outputting the object position coordinates (Col. 5, Line 43 to Col. 6, line 3, Col. 11, Line 29 to Col. 12, Line 6).

Regarding Claim 33, Gordon et al. teaches determining the presence of a user from the clusters by checking features within a presence detection region (Col. 5, Line 43 to Col. 6, line 3, Col. 11, Line 29 to Col. 12, Line 6).

Regarding Claim 34, Gordon et al. teaches calculating the position for each of the clusters excludes those features in the clusters that are outside of an object detection region (Col. 5, Line 43 to Col. 6, line 3, Col. 11, Line 29 to Col. 12, Line 6).

Regarding Claim 35, Onda teaches defining a dynamic object detection region based on the object position coordinates (Col. 1, Lines 12-19).

Regarding Claim 36, Pryor et al. teaches the dynamic object detection region is defined relative to a user's body (page 10, paragraph 244).

Regarding Claim 37, Pryor et al. teaches defining a body position detection region based

on the object position coordinates (page 30, paragraph 583).

Regarding Claim 38, Pryor et al. teaches the body position detection region further

includes detecting a head position of the user (page 14, paragraph 308).

Regarding Claim 39, Pryor et al. teaches smoothing the motion of the object position

coordinates to eliminate jitter between consecutive image frames (page 3, paragraphs 53,54).

Regarding Claim 40, Pryor et al. teaches calculating hand orientation information from

the object position coordinates (page 5, paragraph 99, page 10, paragraphs 231-234 and 238-

241).

Regarding Claim 41, Pryor et al. teaches outputting the object position coordinates

further includes outputting the hand orientation information (page 5, paragraph 99, page 10,

paragraphs 231-234 and 238-241).

Regarding Claim 42, page 3, paragraph 53,54teaches smoothing the changes in the hand

orientation information (page 3, paragraphs 53,54).

Regarding Claim 43, Pryor et al. teaches defining the dynamic object detection region

includes: identifying a position of a torso-divisioning plane from the collection of features; and

determining the position of a hand detection region relative to the torso-divisioning plane in the

axis perpendicular to the torso divisioning plane (page 3, paragraphs 53,54, page 5, paragraphs

99, 105-115, page 6, paragraph 116, page 9, paragraphs 216-224, page 10, paragraphs 230-234

and 238-241).

Regarding Claim 44, Pryor et al. teaches identifying a body center position and a body

boundary position from the collection of features; identifying a position indicating part of an arm

of the user from the collection of features using the intersection of the feature pair cluster with

the torso divisioning plane; and identifying the arm as either a left arm or a right arm using the

arm position relative to the body position (page 3, paragraphs 53,54, page 5, paragraphs 99, 105-

115, page 6, paragraph 116, page 9, paragraphs 216-224, page 10, paragraphs 230-234 and 238-

241).

Regarding Claim 45, Pryor et al. teaches establishing a shoulder position from the body

center position, the body boundary position, the torso-divisioning plane, and the left arm or the

right arm identification (page 3, paragraphs 53,54, page 5, paragraphs 99, 105-115, page 6,

paragraph 116, page 9, paragraphs 216-224, page 10, paragraphs 230-234 and 238-241).

Regarding Claim 46, Pryor et al. teaches the dynamic object detection region includes

determining position data for the hand detection region relative to the shoulder position (page 3,

paragraphs 53,54, page 5, paragraph 99, page 10, paragraphs 231-234 and 238-241).

Regarding Claim 47, Pryor et al. teaches smoothing the position data for the hand

detection region (page 3, paragraphs 53,54, page 5, paragraph 99, page 10, paragraphs 231-234

and 238-241).

Regarding Claim 48, Pryor et al. teaches determining the position of the dynamic object

detection region relative to the torso divisioning plane in the axis perpendicular to the torso

divisioning plane; determining the position of the dynamic object detection region in the

horizontal axis relative to the shoulder position; and determining the position of the dynamic

object detection region in the vertical axis relative to an overall height of the user using the body

boundary position (page 3, paragraphs 53,54, page 5, paragraphs 99, 105-115, page 6, paragraph

116, page 9, paragraphs 216-224, page 10, paragraphs 230-234 and 238-241).

Regarding Claim 49, Pryor et al. teaches the dynamic object detection region includes:

establishing the position of a top of the user's head using topmost feature pairs of the collection

of features unless the topmost feature pairs are at the boundary; and determining the position of a

hand detection region relative to the top of the user's head (page 3, paragraphs 53,54, page 5,

paragraphs 99, 105-115, page 6, paragraph 116, page 9, paragraphs 216-224, page 10, paragraphs

230-234 and 238-241).

Regarding Claim 50, Pryor et al. teaches a method of using stereo vision to interface with

a computer (page 1, paragraph 3), the method comprising: capturing a stereo image using a

stereo camera (page 8, paragraph 173); processing the stereo image to determine position

information of an object in the stereo image (page 10, paragraph 239), the object being

controlled by a user (page 10, paragraph 239); processing the stereo image to identify feature

information, to produce a scene description from the feature information (page 11, paragraph

247), and to identify matching pairs of features in the stereo image (page 5, paragraphs 114,115);

calculating a disparity and a position for each matching feature pair to create the scene

description (page 13, paragraphs 293-297, page 4, paragraph 79); analyzing the scene description

in a scene analysis process to determine position information of the object; clustering the feature

information in a region of interest into clusters having a collection of features by comparison to

neighboring feature information within a predefined range (page 5, paragraphs 111-115, page 6,

paragraphs 116-124)); calculating a position for each of the clusters; and using the position

information allow the user to interact with a computer application (page 10, paragraphs 238-

244).

However, Pryor et al. fails to teach specifically eliminating those clusters having less than

a predefined threshold of features; selecting the position of the clusters that match a predefined

criteria; recording the position of the clusters that match the predefined criteria as object position

coordinates; and outputting the object position coordinates; determining the presence of a user

from the clusters by checking features within a presence detection region and calculating the

position for each of the clusters excludes those features in the clusters that are outside of an

object detection region.

However, Gordon et al. teaches eliminating those clusters having less than a predefined

threshold of features (Col. 5, Line 43 to Col. 6, line 3, Col. 11, Line 29 to Col. 12, Line 6);

selecting the position of the clusters that match a predefined criteria; recording the position of the

clusters that match the predefined criteria as object position coordinates; and outputting the

object position coordinates (Col. 5, Line 43 to Col. 6, line 3, Col. 11, Line 29 to Col. 12, Line 6);

determining the presence of a user from the clusters by checking features within a presence

detection region (Col. 5, Line 43 to Col. 6, line 3, Col. 11, Line 29 to Col. 12, Line 6) and

calculating the position for each of the clusters excludes those features in the clusters that are

outside of an object detection region (Col. 5, Line 43 to Col. 6, line 3, Col. 11, Line 29 to Col.

12, Line 6).

Thus it would have been obvious to one in the ordinary skill in the art at the time of

invention was made to incorporate the teaching of Gordon et al. in Pryor et al. teaching, to be

able to provide computer and a technique for automatically distinguishing between a

background scene and foreground objects in an image.

Prior et al. teaches capturing the stereo image further includes capturing a reference

image from a reference camera and a comparison image from a comparison camera; and

processing the stereo image further includes processing the reference image and the comparison

image to create pairs of features (page 6, paragraph 117, Lines 1-4, paragraph 119, paragraph

121, 128, page 7, paragraph 134, Lines 3-5, paragraph 135, 136, 148-160, page 10, paragraph

238,239).

However, Prior et al. fails to specifically teach the pixel blocks are 8.times.8 pixel

blocks; analyzing the scene description to determine the position information of the object

further includes cropping the scene description to exclude feature information lying outside of a

region of interest in a field of view analyzing the scene description to determine the position

information of the object further includes: clustering the feature information in a region of

interest into clusters having a collection of features by comparison to neighboring feature

information within a predefined range; and calculating a position for each of the clusters

eliminating those clusters having less than a predefined threshold of features; defining a dynamic

object detection region based on the object position coordinates.

However, Onda teaches the pixel blocks are 8.times.8 pixel blocks (Col. 13, Line 60 to

Col. 14, Line 28); analyzing the scene description to determine the position information of the

object further includes cropping the scene description to exclude feature information lying

outside of a region of interest in a field of view (Col. 11, Lines 9-25, Lines 43-58); analyzing the

scene description to determine the position information of the object further includes: clustering

the feature information in a region of interest into clusters having a collection of features by

comparison to neighboring feature information within a predefined range; and calculating a

position for each of the clusters (Col. 11, Lines 9-25, Lines 43-58, Col. 13, Line 60 to Col. 14,

Line 59); eliminating those clusters having less than a predefined threshold of features (page 10,

paragraph 234, Lines 1-4, paragraph 238) and defining a dynamic object detection region based

on the object position coordinates (Col. 1, Lines 12-19).

Thus it would have been obvious to one in the ordinary skill in the art at the time of

invention was made to incorporate the teaching of Onda in Pryor et al. teaching, to be able to

provide computer and a technique for matching stereo images and a method of detecting

disparity between these images to detect positional information in the image pickup space based

on stereo images, volume compression of overall stereo images display control of these stereo

images and for the optical flow extraction of moving images.

Regarding Claim 51, Pryor et al. teaches mapping the position of the object from the feature information from camera coordinates to screen coordinates associated with the computer application; and using the mapped position to interface with the computer application (page 11, paragraph 247-249).

Regarding Claim 52, Pryor et al. teaches recognizing a gesture associated with the object by analyzing changes in the position information of the object in the scene description; and combining the position information and the gesture to interface with the computer application (page 11, paragraph 247-249).

Regarding Claim 53, Pryor et al. teaches the step of capturing the stereo image further includes capturing the stereo image using a stereo camera (page 5, paragraph 111-115).

6.   Claims 71-91,96-98 are rejected under 35 U.S.C. 103(a) as being unpatentable over Pryor et al. (US 2004/0046736 A1) as applied to claims 1-12,23,54-70, above, and further in view of Maurer et al. (US 2001/0033675 A1).

Regarding Claim 71, Pryor et al. teaches a stereo vision system (page 1, paragraph 3, Lines 5-7, paragraph 28, 11-13) for interfacing with an application program running on a computer (page 1, paragraph 3, Lines 5-7, page 5, paragraph 111, Line 2) the stereo vision system (page 1, paragraph 3, Lines 5-7, paragraph 28, 11-13) comprising: first and second video cameras (page 5, paragraph 111, Lines 2,3) arranged in an adjacent configuration and operable to

produce a series of stereo video images (page 6, paragraph 119, Lines 1-7, figure 1c); and a

processor operable to receive the series of stereo video images and detect objects appearing in an

intersecting field of view of the cameras (page 6, paragraphs 119-121), the processor executing a

process to: define an object detection region in three-dimensional coordinates relative to a

position of the first and second video cameras (page 11, paragraphs 247, 249); select a control

object appearing within the object detection region; and map position coordinates of the control

object to a position indicator associated with the application program as the control object moves

within the object detection region (page 11, paragraphs 247-249) and identify an object

perceived as the largest object appearing in the intersecting field of view of the cameras (page

10, paragraphs 243,244) and positioned at a predetermined depth range (page 13, paragraphs

297, page 18, paragraph 383); select the object as an object of interest; determine a position

coordinate representing a position of the object of interest; and use the position coordinate as an

object control point to control the application program (page 18, paragraphs 382, 385, 386,387).

However, Prior et al. fails to specifically teach the indicator is an avatar

However, Maurer et al. teaches the indicator is an avatar (page 5, paragraph 56, (page 6,

paragraphs 66-69, page 7, paragraphs 71-73).

Thus it would have been obvious to one in the ordinary skill in the art at the time of

invention was made to incorporate the teaching of Maurer et al. in Pryor et al. teaching, to be

able to provide a dynamic facial feature sensing, and more particularly, to a vision-based motion

capture system that allows real-time finding, tracking and classification of facial features for

input into a graphics engine that animates an avatar.

Regarding Claim 72, Pryor et al. teaches the process causes the processor to: determine

and store a neutral control point position; map a coordinate of the object control point relative to

the neutral control point position; and use the mapped object control point coordinate to control

the application program (page 11, paragraphs 246-249).

Regarding Claim 73, Pryor et al. teaches the process causes the processor to: define a

region having a position based upon the position of the neutral control point position; map the

object control point relative to its position within the region; and use the mapped object control

point coordinate to control the application program (page 11, paragraphs 246-255).

Regarding Claim 74, Pryor et al. teaches the process causes the processor to: transform

the mapped object control point to a velocity function; determine a viewpoint associated with a

virtual environment of the application program; and use the velocity function to move the

viewpoint within the virtual environment (page 13, paragraph 297, page 14, paragraph 298-310).

Regarding Claim 75, Pryor et al. teaches the process causes the processor to map a

coordinate of the object control point to control a position of an indicator within the application

program (page 13, paragraph 297, page 14, paragraph 298-310, page 11, paragraphs 246-255).

Regarding Claim 76, Pryor et al. teaches capturing the stereo image further includes

capturing a reference image from a reference camera and a comparison image from a comparison

camera; and processing the stereo image further includes processing the reference image and the

comparison image to create pairs of features (page 6, paragraph 117, Lines 1-4, paragraph 119,

paragraph 121, 128, page 7, paragraph 134, Lines 3-5, paragraph 135, 136, 148-160, page 10,

paragraph 238,239).

However, Prior et al. fails to specifically teach the indicator is an avatar

However, Maurer et al. teaches the indicator is an avatar (page 5, paragraph 56, (page 6,

paragraphs 66-69, page 7, paragraphs 71-73).

Thus it would have been obvious to one in the ordinary skill in the art at the time of

invention was made to incorporate the teaching of Maurer et al. in Pryor et al. teaching, to be

able to provide a dynamic facial feature sensing, and more particularly, to a vision-based motion

capture system that allows real-time finding, tracking and classification of facial features for

input into a graphics engine that animates an avatar.

Regarding Claim 77, Maurer et al. teaches the process causes the processor to map a

coordinate of the object control point to control an appearance of an indicator within the

application program (page 6, paragraphs 66-69, page 7, paragraphs71-73, 80-83).

Regarding Claim 78, Maurer et al. teaches the indicator is an avatar (page 6, paragraphs

66-69, page 7, paragraphs71-73)

Regarding Claim 79, Maurer et al. teaches the object of interest is a human appearing

within the intersecting field of view (page 6, paragraphs 66-69, page 7, paragraphs71-73, 80-83).

Regarding Claim 80, Pryor et al. teaches a stereo vision system (page 1, paragraph 3,

Lines 5-7, paragraph 28, 11-13) for interfacing with an application program running on a

computer (page 1, paragraph 3, Lines 5-7, page 5, paragraph 111, Line 2) the stereo vision

system (page 1, paragraph 3, Lines 5-7, paragraph 28, 11-13) comprising: first and second video

cameras (page 5, paragraph 111, Lines 2,3) arranged in an adjacent configuration and operable to

produce a series of stereo video images (page 6, paragraph 119, Lines 1-7, figure 1c); and a

processor operable to receive the series of stereo video images and detect objects appearing in an

intersecting field of view of the cameras (page 6, paragraphs 119-121), the processor executing a

process to: define an object detection region in three-dimensional coordinates relative to a

position of the first and second video cameras (page 11, paragraphs 247, 249); select a control

object appearing within the object detection region; and map position coordinates of the control

object to a position indicator associated with the application-program as the control object moves

within the object detection region (page 11, paragraphs 247-249) and identify an object

perceived as the largest object appearing in the intersecting field of view of the cameras (page

10, paragraphs 243,244) and positioned at a predetermined depth range (page 13, paragraphs

297, page 18, paragraph 383); select the object as an object of interest; determine a position

coordinate representing a position of the object of interest; and use the position coordinate as an

object control point to control the application program (page 18, paragraphs 382, 385, 386,387)

and the control region being positioned at a predetermined location (page 18, paragraph 387)

and having a predetermined size relative to a size and a location of the object of interest; search

the control region for a point associated with the object of interest that is closest to the cameras

(page 18, paragraph 387, Lines 1-6) and within the control region; select the point associated

with the object of interest as a control point if the point associated with the object of interest is

within the control region (page 18, paragraph 387, 6-12); and map position coordinates of the

control point, as the control point moves within the control region, to a position indicator

associated with the application program (page 18, paragraph 387, Lines 1-12, paragraph 391).

However, Prior et al. fails to specifically teach the object of interest is a human

appearing within the intersecting field of view.

However, Maurer et al. teaches the object of interest is a human appearing within the

intersecting field of view (page 6, paragraphs 66-69, page 7, paragraphs71-73, 80-83).

Thus it would have been obvious to one in the ordinary skill in the art at the time of

invention was made to incorporate the teaching of Maurer et al. in Pryor et al. teaching, to be

able to provide a dynamic facial feature sensing, and more particularly, to a vision-based motion

capture system that allows real-time finding, tracking and classification of facial features for

input into a graphics engine that animates an avatar.

Regarding Claim 81, Pryor et al. teaches the processor is operable to: map a horizontal

position of the control point relative to the video cameras to an x-axis screen coordinate of the

position indicator; map a vertical position of the control point relative to the video cameras to a

y-axis screen coordinate of the position indicator; and emulate a mouse function using a

combination of the x-axis and the y-axis screen coordinates (page 10, paragraphs 229-242, page

13, paragraph 297).

Regarding Claim 82, Pryor et al. teaches the processor is operable to: map a x-axis

position of the control point relative to the video cameras to an x-axis screen coordinate of the

position indicator; map a y-axis position of the control point relative to the video cameras to a y-

axis screen coordinate of the position indicator; and map a z-axis depth position of the control

point relative to the video cameras to a virtual z-axis screen coordinate of the position indicator

(page 13, paragraph 297, page 14, paragraph 298-310, page 11, paragraphs 246-255).

Regarding Claim 83, Pryor et al. teaches capturing the stereo image further includes

capturing a reference image from a reference camera and a comparison image from a comparison

camera; and processing the stereo image further includes processing the reference image and the

comparison image to create pairs of features (page 6, paragraph 117, Lines 1-4, paragraph 119,

paragraph 121, 128, page 7, paragraph 134, Lines 3-5, paragraph 135, 136, 148-160, page 10,

paragraph 238,239).

However, Prior et al. fails to specifically teach the object of interest is a human

appearing within the intersecting field of view.

However, Maurer et al. teaches the object of interest is a human appearing within the

intersecting field of view (page 6, paragraphs 66-69, page 7, paragraphs71-73, 80-83).

Thus it would have been obvious to one in the ordinary skill in the art at the time of

invention was made to incorporate the teaching of Maurer et al. in Pryor et al. teaching, to be

able to provide a dynamic facial feature sensing, and more particularly, to a vision-based motion

capture system that allows real-time finding, tracking and classification of facial features for

input into a graphics engine that animates an avatar.

Regarding Claim 84, Prior et al. teaches the control point is associated with a human

hand appearing within the control region (page 10, paragraphs 230-238).

Regarding Claim 85, Pryor et al. teaches a stereo vision system (page 1, paragraph 3,

Lines 5-7, paragraph 28, 11-13) for interfacing with an application program running on a

computer (page 1, paragraph 3, Lines 5-7, page 5, paragraph 111, Line 2) the stereo vision

system (page 1, paragraph 3, Lines 5-7, paragraph 28, 11-13) comprising: first and second video

cameras (page 5, paragraph 111, Lines 2,3) arranged in an adjacent configuration and operable to

produce a series of stereo video images (page 6, paragraph 119, Lines 1-7, figure 1c); and a

processor operable to receive the series of stereo video images and detect objects appearing in an

intersecting field of view of the cameras (page 6, paragraphs 119-121), the processor executing a

process to: define an object detection region in three-dimensional coordinates relative to a

position of the first and second video cameras (page 11, paragraphs 247, 249); select a control

object appearing within the object detection region; and map position coordinates of the control

object to a position indicator associated with the application program as the control object moves

within the object detection region (page 11, paragraphs 247-249) and identify an object

perceived as the largest object appearing in the intersecting field of view of the cameras (page

10, paragraphs 243,244) and positioned at a predetermined depth range (page 13, paragraphs

297, page 18, paragraph 383); select the object as an object of interest; determine a position

coordinate representing a position of the object of interest; and use the position coordinate as an

object control point to control the application program (page 18, paragraphs 382, 385, 386,387)

and as the hand objects move within the object detection region, to positions of virtual hands

associated with an avatar rendered by the application program (page 11, paragraphs 251-257,

hand motion generates virtual mouse operation or as appointing device or paint brush).

However, Prior et al. fails to specifically teach the avatar takes the form of a human-like

body.

However, Maurer et al. teaches the avatar takes the form of a human-like body (page 6,

paragraphs 66-69, page 7, paragraphs71-73, 80-83).

Thus it would have been obvious to one in the ordinary skill in the art at the time of

invention was made to incorporate the teaching of Maurer et al. in Pryor et al. teaching, to be

able to provide a dynamic facial feature sensing, and more particularly, to a vision-based motion

capture system that allows real-time finding, tracking and classification of facial features for

input into a graphics engine that animates an avatar.

Regarding Claim 86, Prior et al. teaches the process selects the up to two hand objects

from the objects appearing in the intersecting field of view that are closest to the video cameras

and within the object detection region (page 10, paragraphs 230-238).

Regarding Claim 87, Pryor et al. teaches capturing the stereo image further includes

capturing a reference image from a reference camera and a comparison image from a comparison

camera; and processing the stereo image further includes processing the reference image and the

comparison image to create pairs of features (page 6, paragraph 117, Lines 1-4, paragraph 119,

paragraph 121, 128, page 7, paragraph 134, Lines 3-5, paragraph 135, 136, 148-160, page 10, paragraph 238,239).

However, Prior et al. fails to specifically teach the avatar takes the form of a human-like body.

However, Maurer et al. teaches the avatar takes the form of a human-like body (page 6, paragraphs 66-69, page 7, paragraphs71-73, 80-83).

Thus it would have been obvious to one in the ordinary skill in the art at the time of invention was made to incorporate the teaching of Maurer et al. in Pryor et al. teaching, to be able to provide a dynamic facial feature sensing, and more particularly, to a vision-based motion capture system that allows real-time finding, tracking and classification of facial features for input into a graphics engine that animates an avatar.

Regarding Claim 88, Maurer et al. teaches the avatar is rendered in and interacts with a virtual environment forming part of the application program (page 1, paragraph 2).

Regarding Claim 89, Maurer et al. teaches the processor further executes a process to compare the positions of the virtual hands associated with the avatar to positions of virtual objects within the virtual environment to enable a user to interact with the virtual objects within the virtual environment (page 1, paragraph 2).

Prior et al. teaches the processor further executes a process to compare the positions of the virtual hands associated with positions of virtual objects within the virtual environment to

enable a user to interact with the virtual objects within the virtual environment (page 10,

paragraphs 230-243).

Regarding Claim 90, Prior et al. teaches the processor further executes a process to:

detect position coordinates of a user within the intersecting field of view (page 10, paragraphs

230-243); and Maurer et al. teaches map the position coordinates of the user to a virtual torso of

the avatar rendered by the application program (page 1, paragraph 2, page 6, paragraphs 66-69,

page 7, paragraphs71-73, 80-83).

Regarding Claim 91, Prior et al. the process moves at least one of the virtual hands to a

neutral position if a corresponding hand object is not selected (page 12, paragraphs 263-270).

Maurer et al. teaches the avatar is rendered in and interacts with a virtual environment

forming part of the application program (page 1, paragraph 2).

Regarding Claim 96, Maurer et al. teaches a virtual knee position associated with the

avatar is derived by the application program and used to refine an appearance of the avatar (page

1, paragraph 2, page 6, paragraphs 66-69, page 7, paragraphs71-73, 80-83, since it can work with

face and facial features it can work with elbow or knee).

Regarding Claim 97, Maurer et al. teaches a virtual elbow position associated with the

avatar is derived by the application program and used to refine an appearance of the avatar (page

1, paragraph 2, page 6, paragraphs 66-69, page 7, paragraphs71-73, 80-83, since it can work with

face and facial features it can work with elbow or knee).

Regarding Claim 98, Pryor et al. teaches a third video camera arranged in an adjacent

configuration with the first and second video cameras and operable to produce the series of

stereo video images (page 1, paragraph 3, Lines 4-7).

Allowable Subject Matter

7.   Claims 92-95 objected to as being dependent upon a rejected base claim, but would be

allowable if rewritten in independent form including all of the limitations of the base claim and

any intervening claims.

8.   The following is an examiner's statement of reasons for allowance:

a stereo vision system for interfacing with an application program running on a computer,

the stereo vision system comprising: first and second video cameras arranged in an adjacent

configuration and operable to produce a series of stereo video images; and a processor operable

to receive the series of stereo video images and detect objects appearing in an intersecting field

of view of the cameras, the processor executing a process to: define an object detection region in

three-dimensional coordinates relative to a position of the first and second video cameras; select

up to two hand objects from the objects appearing in the intersecting field of view that are within

the object detection region; and map position coordinates of the hand objects, as the hand objects

move within the object detection region, to positions of virtual hands associated with an avatar

rendered by the application program and **the processor further executes a process to: detect**

position coordinates of a user within the intersecting field of view; and map the position

coordinates of the user to a velocity function that is applied to the avatar to enable the

avatar to roam through a virtual environment rendered by the application program and

the velocity function includes a neutral position denoting zero velocity of the avatar; map

the position coordinates of the user relative to the neutral position into torso coordinates

associated with the avatar so that the avatar appears to lean and compare the position of

the virtual hands associated with the avatar to positions of virtual objects within the virtual

environment to enable the user to interact with the virtual objects while roaming through

the virtual environment.

The cited references of 892's fail to anticipate individually or render obviousness

individually as well as in combination the underlined above.

Any comments considered necessary by applicant must be submitted no later than the

payment of the issue fee and, to avoid processing delays, should preferably accompany the issue

fee. Such submissions should be clearly labeled "Comments on Statement of Reasons for

Allowance."

9.   The prior art made of record and not relied upon is considered pertinent to applicant's

disclosure. Applicant is informed that all of the other additional cited references either anticipate

or render the claims obvious. In order to not to be repetitive and exhaustive, the examiner did

draft additional rejection based on those references.

Conclusion

10.   The prior art made of record and not relied upon is considered pertinent to applicant's

disclosure.

   Sudarsky et al. (6,088,035) Method for displaying graphic model.

11.   Any inquiry concerning this communication or earlier communications from the

examiner should be directed to Prabodh M Dharia whose telephone number is 703-605-1231.

The examiner can normally be reached on M-F 8AM to 5PM.

12.   If attempts to reach the examiner by telephone are unsuccessful, the examiner's

supervisor, Bipin Shalwala can be reached on 703-3054938. The fax phone number for the

organization where this application or proceeding is assigned is 703-872-9306.

13.   Information regarding the status of an application may be obtained from the Patent

Application Information Retrieval (PAIR) system. Status information for published applications

may be obtained from either Private PAIR or Public PAIR. Status information for unpublished

applications is available through Private PAIR only. For more information about the PAIR

system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR

system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free).
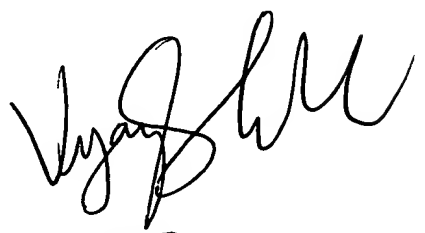
   Any response to this action should be mailed to:

   Commissioner of Patents and Trademarks

   Washington, D.C. 20231

PD

AU2673

VIJAY SHANKAR
PRIMARY EXAMINER